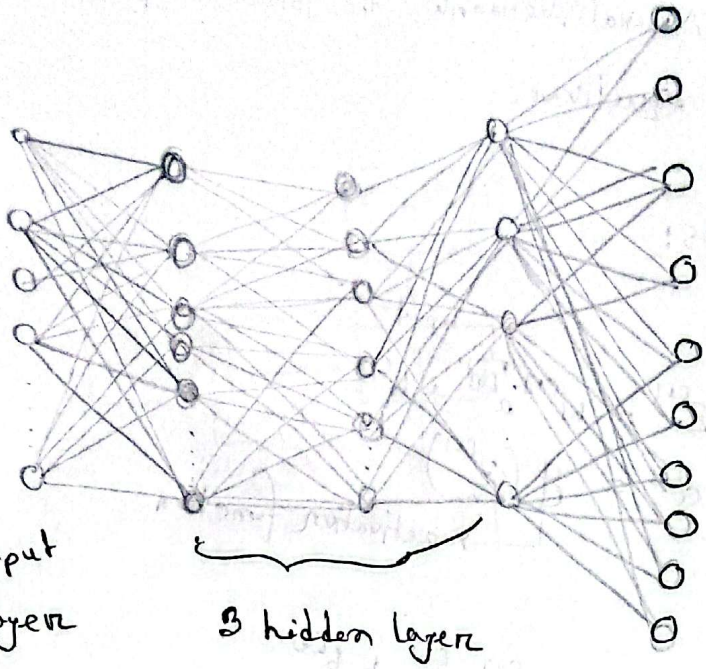
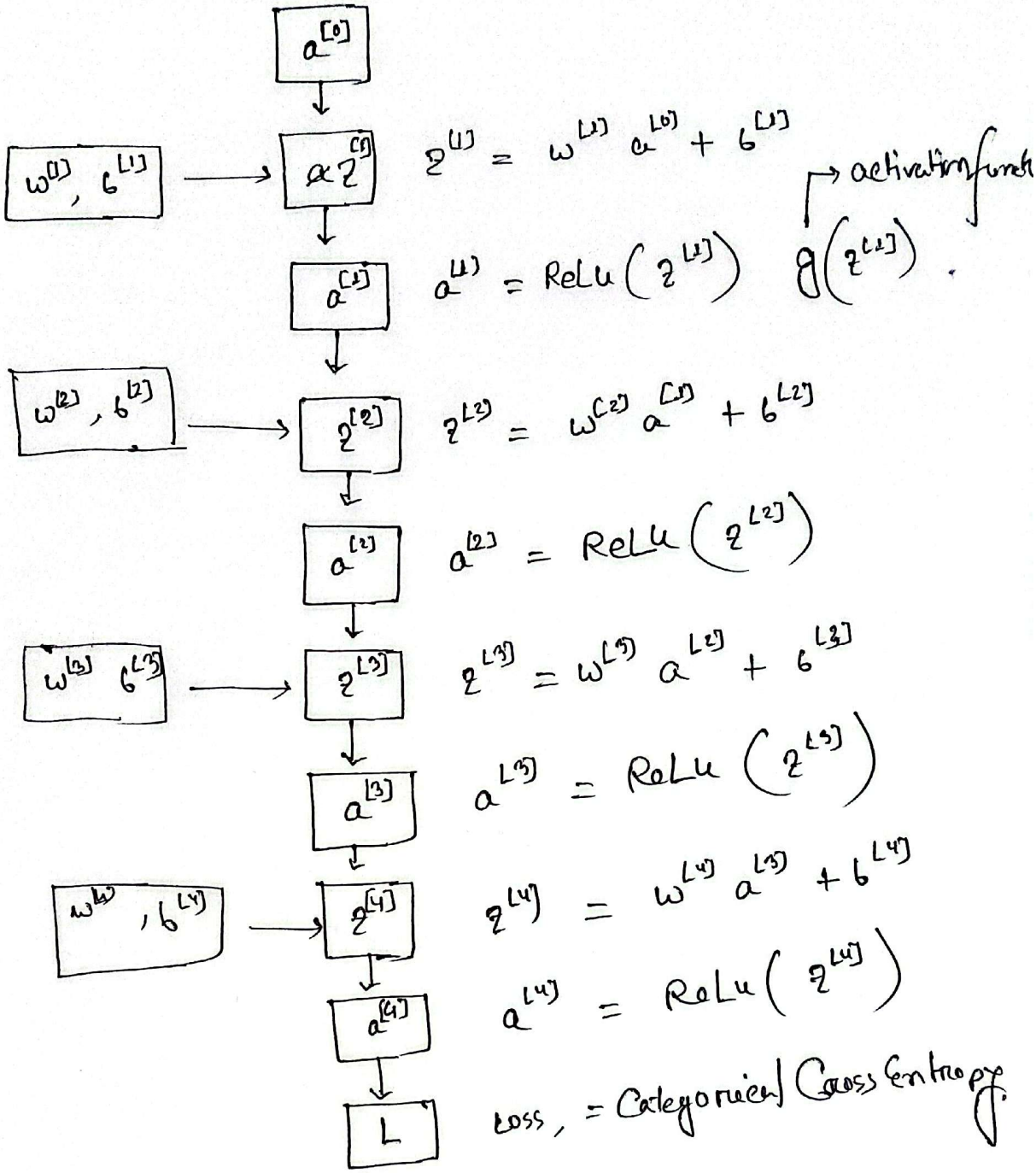
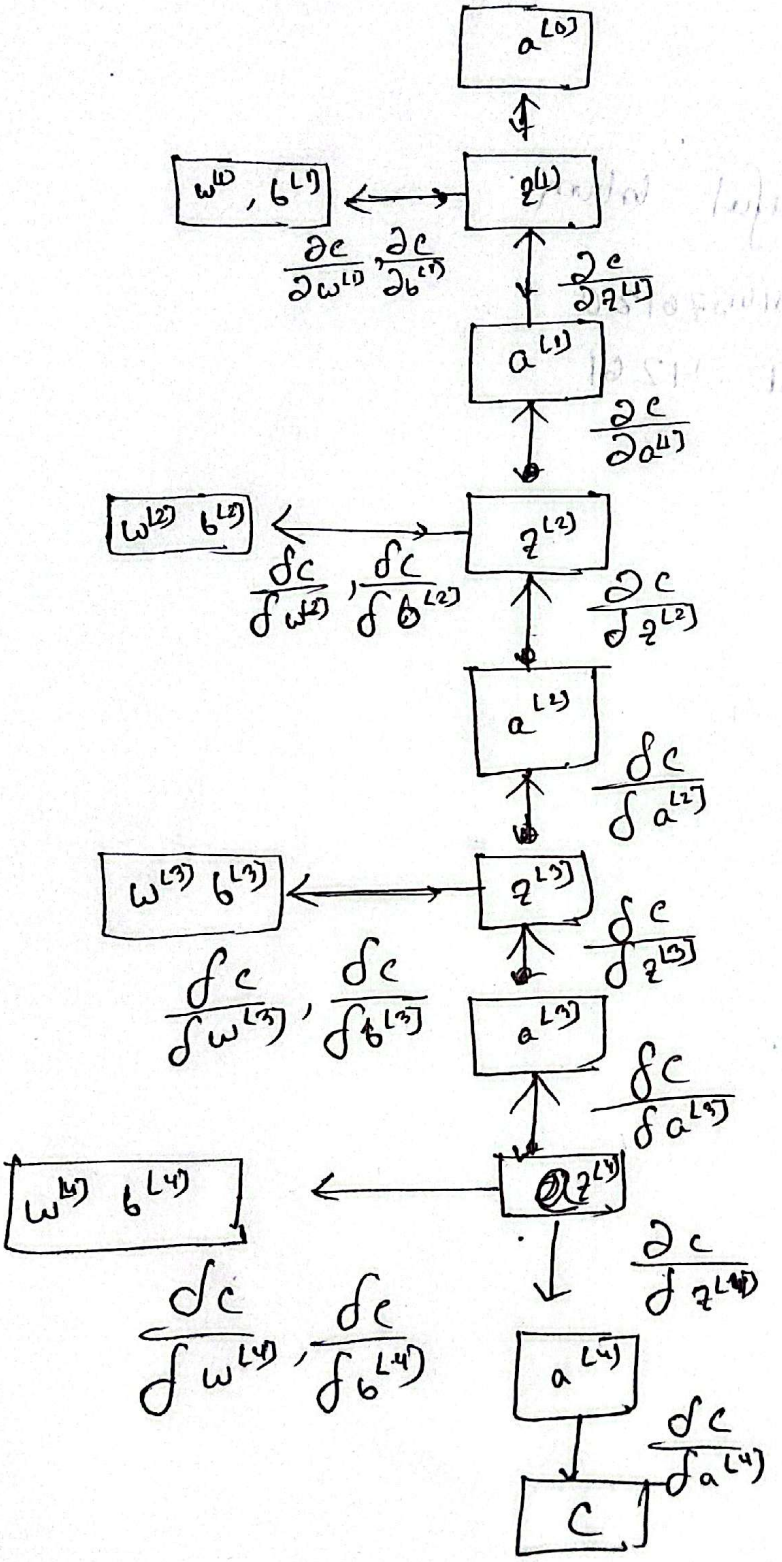
1.



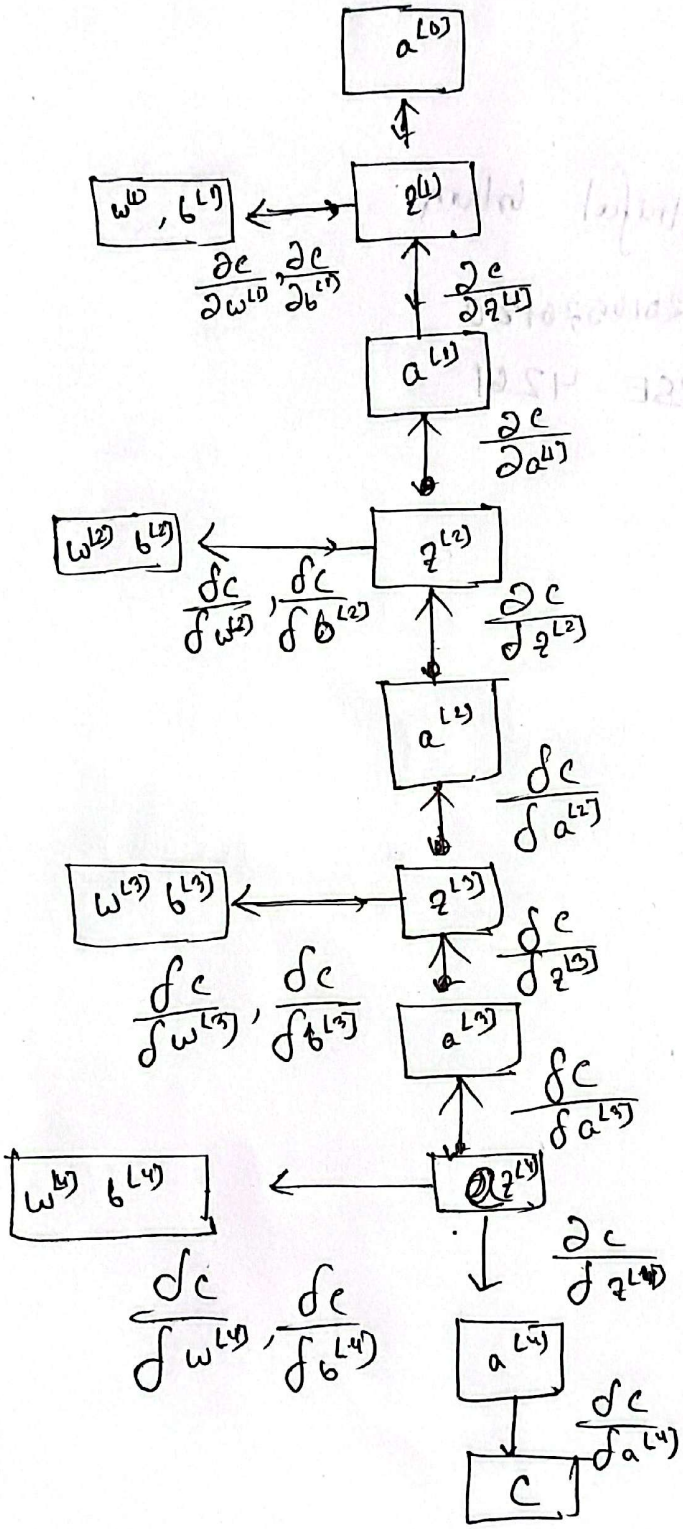
⇒ forword propagation Computation Graph.



3) Backpropagation Computation Graph,



Backpropagation Computation Graph,



* For 5 layer Neural Network, the forword pass and backword pass gradient equations.

Forword pass:

layer 1:

$$z^{[1]} = w^{[1]} a^{[0]} + b^{[1]}$$

($a^{[0]}$: input)

$$a^{[1]} = g(z^{[1]})$$

(g: activation function)

layer 2:

$$z^{[2]} = w^{[2]} a^{[1]} + b^{[2]}$$

$$a^{[2]} = g(z^{[2]})$$

Layer 3

$$z^{[3]} = w^{[3]} a^{[2]} + b^{[3]}$$

$$a^{[3]} = g.(z^{[3]})$$

layer 4:

$$z^{[4]} = w^{[4]} a^{[3]} + b^{[4]}$$

$$a^{[4]} = g.(z^{[4]})$$

layer 5:

$$z^{5} = w^{[5]}.a^{[4]} + b^{[4]}$$

$$a^{[5]} = g.(z^{[5]})$$

Backward propagation:

Layer 5:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta w^{[4]}}$$

(Cross entropy loss function)

Layer 4:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta w^{[3]}}$$

Layer 3:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta a^{[2]}} \cdot \frac{\delta a^{[2]}}{\delta z^{[2]}} \cdot \frac{\delta z^{[2]}}{\delta w^{[2]}}$$

Layer 2:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{4}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta a^{[2]}} \cdot \frac{\delta a^{[2]}}{\delta z^{[2]}} \cdot \frac{\delta z^{[2]}}{\delta a^{[1]}} \cdot \frac{\delta a^{[1]}}{\delta z^{[1]}} \cdot \frac{\delta z^{[1]}}{\delta w^{[1]}}$$

Backward propagation:

Layer 5:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta w^{[4]}}$$

→ Cross entropy loss function

Layer 4:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta w^{[3]}}$$

Layer 3:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{[4]}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta a^{[2]}} \cdot \frac{\delta a^{[2]}}{\delta z^{[2]}} \cdot \frac{\delta z^{[2]}}{\delta w^{[2]}}$$

Layer 2:

$$\frac{\delta C}{\delta w} = \frac{\delta C}{\delta a^{4}} \cdot \frac{\delta a^{[4]}}{\delta z^{[4]}} \cdot \frac{\delta z^{[4]}}{\delta a^{[3]}} \cdot \frac{\delta a^{[3]}}{\delta z^{[3]}} \cdot \frac{\delta z^{[3]}}{\delta a^{[2]}} \cdot \frac{\delta a^{[2]}}{\delta z^{[2]}} \cdot \frac{\delta z^{[2]}}{\delta a^{[1]}} \cdot \frac{\delta a^{[1]}}{\delta z^{[1]}} \cdot \frac{\delta z^{[1]}}{\delta w^{[1]}}$$

6)

We train same neural network model by Manual training with "tf.GradientTape" and "model.fit()".

In both case, the parameters are,

epoch → 5
Batch size → 128
Loss → Categorical Cross Entropy
Optimizer → Adam.

the testset Accuracy in tf.GradientTape = 0.9644.

the test set " " model.fit() = 0.9673